

# Capstone Project Proposal

## Problem Statement

The proposal is to investigate the feasibility of and construct a predictive financial model, capable of estimating the future price movement of traded financial instruments based primarily on historical price data.

This is a challenging project, since there are a range of views held as to whether this is feasible. On one hand, the Efficient Market Hypothesis (EMH) [1] argues that for efficient markets, it should not be possible to predict future prices, based on historic price information alone. On the other hand, investing strategies which rely on exactly this approach, such as momentum and trend-following approaches [2] do so with empirical success under certain conditions, suggesting that there may indeed be structure in historic price information which can be used to predict future price movements.

There are many factors that influence the price of a financial product traded on any particular market at a given time. These range from the sentiment and demand for the product by the market participants/investors to the dynamics of the systems used to operate the marketplace itself.

The hypothesis proposed here is that some of these factors are not completely independent of recent changes in the price of the product, and that by analysing enough data, some structure may be found in historic price movements that would be a predictor of future price movements.

It is appreciated that any such structure discoverable is likely to be a very weak underlying signal amid a lot of noise. Therefore so it's entirely feasible, indeed likely, that any predictive signal obtained, would not be strong enough to trade profitably on its own. Trading any systematic trading strategy incurs execution costs, typically due to price spreads (the difference between the cost to buy and that to sell a given product) and fees or commissions charged by the various brokers or marketplaces involved.

Nevertheless, it is proposed that even the identification of a weak but measurable signal would be a valuable contribution, since it could be incorporated as an additional input into existing systematic investment strategies to increase their profitability or diversity.

## Project Output

The project output would be a model capable of providing a prediction of the direction of the short-term future price movement of a financial product, along with test results, analysis measurements of its performance on a hold-out test data set unseen during its training.

## Dataset and inputs

To have confidence in any possible discovery of a weak signal a lot of data is required. Hence the direction of the project is driven to some extent by the availability of such data. Whilst there is a wealth of financial data potentially available, most of this comes at significant cost.

However, a source of freely available FX price data was found at [3], which provides open/high/low/close FX prices sampled at 1 minute intervals across many currency pairs.

This led to the proposal to attempt the prediction of short term price movements of FX currency pairs, based on recent price history.

The effect of other inputs, such as temporal indicators, such as hour of day, and day of week may be explored. Certainly FX markets are known to exhibit different volatility at different times of day, as market participants in different time zones become active. Whilst not necessarily contributing to a direct prediction of *direction* of future price moves, it is considered likely that they may have impact on prediction of the distribution of *sizes* of any price moves.

## Proposed Solution

It is proposed to use Deep Neural Networks trained on historical price information to produce a model providing predictions on future pricing movements.

Deep neural networks were chosen as they are known to be capable of capturing arbitrarily complex patterns in high dimensional data, and they may be trained using Stochastic Gradient Descent (and variants thereof) which is a scalable approach to dealing with large volumes of data.

Part of the project will be to determine the relative effectiveness of a regression model, versus a classification model (where future price moves may be categorised into coarse buckets to indicate magnitude and direction, and confidence level).

One tradeoff to be investigated is that between the confidence achievable in very short term prediction with that in a longer term prediction, which may be of lower confidence but of greater price movement.

Some investigation will focus on how a model trained on a single instrument's historic data compares with one trained on data from a range of instruments.

## Benchmark

A number of baseline models are proposed for comparison and benchmarking:

Benchmarks for regression type models:

1. last price predictor. - a model which simply predicts the most recent price for a product as the best estimate of it's future price. This would seem to be a reasonable baseline in keeping with EMH [1], which would infer that no better estimate should be possible.
2. moving average predictor - a model which simply predicts the moving average price over a recent window would be another interesting model for comparison.

The effectiveness of regression models in price prediction could be compared using the mean squared error in predicted price movement versus actual movements over a suitable test data set.

For categorical models metrics such as accuracy, precision/recall and f1 score may be used.

Depending on the nature and distribution of any signal discovered, more directly relevant metrics measuring its value as a trading signal may be developed.

Naturally final evaluation would be on hold-out data sets, unseen during the model development.

## Project Design

Initial investigations will focus on detection of the existence of any patterns of future price movements, conditioned on features in historic data.

Assuming some structure can be discovered, some focus will then be given to the selection of the most effective input features.

Once a range of features have been selected some investigation will be given over to selecting a suitable network architecture, and optimal hyperparameters.

Finally the project will conclude with testing and benchmarking of the selected model, and identification of possible areas for further research.

## **Data preprocessing**

A range of different features based on price history will be considered.

Some of these may be inspired from features often used in momentum strategies, such as ratios of moving averages of recent prices, others may be simply the size of recent price movements over a range of lookback periods.

## **References:**

- [1] [https://en.wikipedia.org/wiki/Efficient-market\\_hypothesis](https://en.wikipedia.org/wiki/Efficient-market_hypothesis) ([https://en.wikipedia.org/wiki/Efficient-market\\_hypothesis](https://en.wikipedia.org/wiki/Efficient-market_hypothesis))
- [2] [https://en.wikipedia.org/wiki/Momentum\\_\(finance\)](https://en.wikipedia.org/wiki/Momentum_(finance)) ([https://en.wikipedia.org/wiki/Momentum\\_\(finance\)](https://en.wikipedia.org/wiki/Momentum_(finance)))
- [3] <http://www.histdata.com/> (<http://www.histdata.com/>)